2614–2620 Nucleic Acids Research, 1995, Vol. 23, No. 14

Cleavage site determinants in the mammalian polyadenylation signal

Fan Chen, Clinton C. MacDonald[1] and Jeffrey Wilusz*

UMDNJ–New Jersey Medical School, Department of Microbiology and Molecular Genetics, 185 South Orange Avenue, Newark, NJ 07103, USA and [1]Department of Molecular Biology, Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544-1014, USA

Received April 28, 1995; Revised and Accepted June 7, 1995

ABSTRACT

Using a series of position and nucleotide variants of the SV40 late polyadenylation signal we have demonstrated that three sequence elements determine the precise site of 3′-end cleavage in mammalian pre-mRNAs: an upstream AAUAAA element, a downstream U-rich element consisting of five nucleotides, at least four of which are uridine, and a nucleotide preference at the site of cleavage in the order A > U > C >> G. Cleavage occurs no closer than 11 bases, but no farther than 23 bases from the AAUAAA element. The downstream U-rich element is usually located 10–30 bases from the cleavage site. The relative position of the AAUAAA and the U-rich elements define the approximate region within a 13 base domain in which cleavage will occur. The exact position of cleavage is then determined by the local nucleotide sequence in the order of preference noted above. This model accounts for nearly three quarters of polyadenylation signals surveyed and is consistent with previous experimental observations.

INTRODUCTION

Maturation of the 3′-end of most mammalian pre-mRNAs involves a site-specific endonucleolytic cleavage event followed by polymerization of 150–200 adenylate residues in a template-independent manner (1). Cleavage and polyadenylation are tightly coupled *in vivo* (2) and can be uncoupled in the *in vitro* system originally described by Moore and Sharp (3). Polyadenylation site choice may influence gene expression through the inclusion or exclusion of specific exons (4). Moreover, polyadenylation signals are essential for transcription termination by RNA polymerase II (5,6) and influence pre-mRNA splicing (7,8).

The hexanucleotide AAUAAA, located upstream of the cleavage/polyadenylation site, is highly conserved among mammalian polyadenylation signals and is absolutely required for 3′-end processing of pre-mRNAs (1). Deletion of the AAUAAA element abolishes cleavage and polyadenylation (9) and point mutations in this sequence greatly reduce the efficiency of 3′-end processing (10,11). Furthermore, the AAUAAA element serves as the binding site for the 160 kDa protein of cleavage polyadenylation specificity factor (CPSF) (12,13). These observations strongly suggest that the AAUAAA element is the major upstream element of the polyadenylation signal.

Since AAUAAA sequences are also found randomly distributed throughout pre-mRNAs, this element alone cannot define a polyadenylation signal. Sequences downstream of the cleavage site have also been shown to play an important role in 3′-end processing. Deletions in the downstream region of many polyadenylation signals decrease the efficiency of 3′-end processing (14–23). Downstream regions are generally GU- and/or U-rich, but defining an adequate consensus sequence which accounts for all experimental observations has proved elusive (18,24,25). We have recently shown that a four-out-of-five base uridylate tract located within 30 bases of the cleavage site could efficiently substitute for the entire downstream region of the SV40 late (SVL) polyadenylation signal (26). Furthermore, this U-rich element (URE) serves as the binding site for the 64 kDa protein of cleavage stimulation factor (CstF) (27–29). A URE, therefore, may be the major downstream element of the polyadenylation signal.

Several studies have suggested the influence of multiple *cis*-acting elements in determining the site of cleavage. Deletions in the downstream region of the *Xenopus* β-globin pre-mRNA were shown to alter the site of cleavage (30). We have also shown that movement of the downstream URE from position +15 to position +20 or +25 in the context of the SVL polyadenylation signal caused a partial shift in cleavage to a location further downstream (26,27). Deletions between the AAUAAA and the wild-type cleavage site of the SVL polyadenylation signal have also been shown to shift the site of cleavage to a more downstream location (9). Finally, it has been observed that the dinucleotide CA is often present at the cleavage site (10), suggesting that it may play a role in cleavage site selection.

In this study we determined the precise rules governing the relationship of the major *cis*-acting elements involved in cleavage site determination. We have characterized the positional requirements for both the AAUAAA element and the U-rich downstream element in 3′-end formation. Using saturation mutagenesis we have also identified sequence preferences for the endonucleolytic cleavage reaction. These data allowed us to

* To whom correspondence should be addressed

develop a model for the sequence and position requirements of the major elements of the mammalian polyadenylation signal, which is followed by almost three quarters of signals surveyed. In addition, this model is well supported by previous analyses of deletions and mutations of polyadenylation signals.

MATERIALS AND METHODS

Plasmids and transcripts

All RNAs were transcribed *in vitro* using SP6 RNA polymerase and [^{32}P]UTP. RNAs were purified from 5% polyacrylamide–7 M urea gels prior to use (31). Positions in the SVL polyadenylation signal are denoted as follows: the wild-type cleavage site is designated as +1; positions downstream of the cleavage site are given positive values and positions upstream of the cleavage site are designated by negative numbers.

Transcripts were derived as follows. pSVL contained the 241 bp *Bam*HI–*Bcl*I fragment of SV40 inserted into the *Bam*HI site of pSP65 (Promega). Transcription of *Dra*I-linearized template yielded a 224 base RNA (previously designated SVL, herein designated SVL-13A). pSVL-sub+6–29, which contained a substitution of the SVL downstream region from position +6 to +29 with vector-derived sequence, has been described previously (26). pSVL-9U was derived by insertion of the synthetic oligonucleotide 5′-TCTTTAAGCTTGGATCCGGAGAGCT-3′ and its appropriate complement between the *Sst*I site and blunted *Cla*I site of pSVL-sub+6–29. pSVL-20U was constructed by inserting the oligonucleotide 5′-CGATAAGCTTGGATCTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTAAACTGCA-3′ and its appropriate complement between the *Bsm*I and *Pst*I sites of pSVL. pSVL-6U, pSVL-14U, pSVL-23U and pSVL-25U, which contain the U-rich element at positions +6, +14, +23 and +25 downstream of the wild-type cleavage site respectively, have been previously designated as pSVL-U6, pSVL-U14, pSVL-U23 and pSVL-25U (26). pSVL-8A, pSVL-20A and their derivatives were generated by a combination of conventional cloning and PCR approaches. A 191 bp region from the beginning of the SP6 promoter to position –10 upstream of the cleavage/polyadenylation site was amplified from the *Sph*I–*Xba*I fragment of pSVL by PCR. The SP6 promoter-specific primer CATACGATTTAGGTGACACTATAG and either primer –8A (GCAATATTTATTTTCTTGCAGCTTATAATGG) or primer –20A (GCAATATTTTGTTTCTTGTTTATTATAATGGTTACAAA) were used to generate constructs in which the AAUAAA element was located either at position –8 or –20 upstream of the cleavage/polyadenylation site. Both primers contained an *Ssp*I site at their 5′-ends. Amplification reactions were performed in 100 µl using standard reaction mixtures for 35 cycles of 94°C (1 min), 49°C (1 min) and 72°C (1 min). Amplified DNAs were purified on 2% low melting point agarose gels prior to use. pSVL-8A and pSVL-8A/23U were prepared by digesting the PCR product generated using primer –8A with *Eco*RI and *Ssp*I and inserting the fragment between the *Eco*RI and *Hpa*I sites of pSVL and pSVL-23U respectively. To construct pSVL-20A, pSVL-20A/6U, pSVL-20A/9U, pSVL-20A/14U, pSVL-20A/20U, pSVL-20A/23U and pSVL-20A/25U, the PCR product generated using primer –20A was cleaved with *Eco*RI and *Ssp*I and inserted between the *Eco*RI and *Hpa*I sites of pSVL, pSVL-6U, pSVL-9U, pSVL-14U, pSVL-20U, pSVL-23U and pSVL-25U respectively.

A set of constructs (designated pSVL-NN), which carry all possible nucleotide combinations at positions –1 and +1 of the SVL polyadenylation signal, was created by insertion of the synthetic oligonucleotide 5′-AACAANNACAATGGGGATCCTTTTTG-3′ (where N refers to an equal mixture of all four nucleotides) and its appropriate complement between the *Hpa*I and *Sal*I sites of either pSVL-14U or pSVL. Individual constructs were identified by sequence analysis. Nucleotide combinations of NN not identified in the first round of screening were synthesized individually.

In vitro cleavage assays

Cleavage reactions were performed using equimolar amounts of the indicated transcripts in the *in vitro* system of Moore and Sharp modified as described previously (31). Nuclear extracts were prepared from HeLa spinner cells grown in 10% horse serum as described (32). A typical 12.5 µl reaction contained a final concentration of 3% polyvinyl alcohol, 1 mM α,β-methylene ATP, 1 mM EDTA, 12 mM HEPES, pH 7.9, 12% glycerol, 60 mM KCl, 0.3 mM dithiothreitol and 60% v/v HeLa nuclear extract. RNA products were analyzed on 5% polyacrylamide sequencing gels containing 8 M urea.

Computer surveys

A total of 52 214 non-viral mammalian genes from the GenBank database (version 81.0, February 1994) were initially searched for sequences which contained eight or more A residues at their 3′-end with an AAUAAA element located 5–40 bases upstream (27). Duplicates and sequences with poly(A/T) were eliminated to yield 1856 sequence entries. The database was searched again using this subset of entries as probes for related sequences that lacked a poly(A) tail. Analysis of the results of this search revealed 131 entries which contained an entire polyadenylation signal with a cleavage site inferred through the sequence of a cDNA clone.

RESULTS

The relative position of the AAUAAA and the URE, as well as a minimal distance from the AAUAAA element, determine the site of cleavage

It has been previously demonstrated that a downstream URE plays a role in the processing of the SV40 late polyadenylation signal (23,26,27,33). Movement of the URE from its normal position 15 bases from the cleavage site (designated +15) to a position beyond 30 bases downstream of the cleavage site resulted in a dramatic reduction in 3′-end processing activity (26). Furthermore, movement of the URE 5–10 bases downstream of its wild-type position caused a partial shift in the site of cleavage to a location three bases downstream (26,27). Movement of the URE closer than position +15 to the cleavage site decreased the efficiency of 3′-end processing, but did not cause a concomitant upstream shift in the site of cleavage (26). These data strongly suggest that the relative position of the URE plays a role in the efficiency of 3′-end processing and in determination of the cleavage site.

We now wished to identify which *cis*-acting element, if any, determines the significance of the relative position of the downstream URE. The most likely candidate was the AAUAAA element, due to its major role in polyadenylation efficiency and the requirement of an AAUAAA element for efficient cross-linking of the URE-specific 64 kDa subunit of CstF to polyadenylation substrate RNAs (31,34). If interactions between the

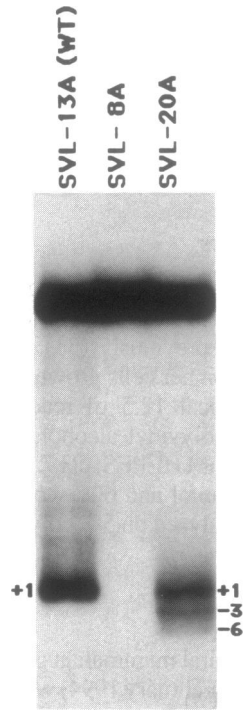

Figure 1. The position of the AAUAAA element influences the efficiency and site of cleavage. A pair of SVL variants which contained the AAUAAA element at position –8 or –20 upstream of the cleavage site were incubated in the *in vitro* cleavage system. RNA products were analyzed on a 5% polyacrylamide gel containing 8 M urea. The 5′ cleavage products at positions +1, –3 and –6 are indicated.

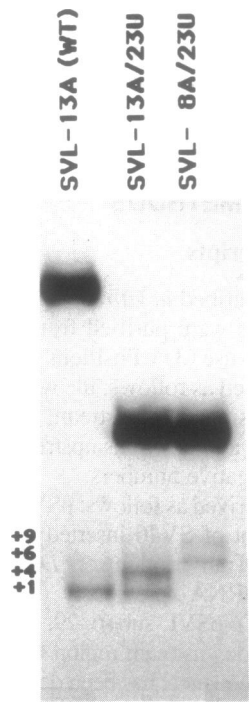

Figure 2. The relative position of the AAUAAA and URE affects the site of cleavage. The URE was moved from position +15 to +23 downstream of the cleavage site in both SVL-13A/23U and SVL-8A/23U and the AAUAAA element was moved from its wild-type position at –13 to –8 upstream of the cleavage site in SVL-8A/23U. These RNAs, along with wild-type SVL-13A RNA, were incubated in the *in vitro* cleavage system and reaction products were analyzed on a 5% polyacrylamide gel containing 8 M urea. The 5′ cleavage products at positions +1, +4, +6 and +9 are indicated.

AAUAAA and U-rich elements are important for 3′-end processing, we predict that movement of the elements in relation to each other would affect the site and efficiency of cleavage. Using a combination of conventional cloning and PCR approaches we moved the AAUAAA and URE to a variety of positions with respect to each other in the SVL polyadenylation signal and tested this hypothesis.

As seen in Figure 1, placement of the AAUAAA element at its wild-type position at –13 caused cleavage to predominantly occur at the position designated +1 [lane SVL-13A(WT)]. Movement of the AAUAAA element five bases downstream to position –8 caused a significant loss of cleavage activity (lane SVL-8A). The lack of efficient cleavage likely results from the relative proximity of the AAUAAA and the URE in this construct (26), although the possibility of inhibitory secondary structures cannot be ruled out. When the AAUAAA element was placed at position –20, seven bases upstream from its wild-type location, 3′-end processing was efficient, but the site of cleavage was partially shifted to locations three and six bases upstream from the wild-type cleavage site (lane SVL-20A). From this data we conclude that, similar to the downstream URE, the position of the AAUAAA in the polyadenylation signal influences both the efficiency and site of cleavage.

We next tested the influence each element had on the positional requirements of the other. As shown in Figure 2, movement of the URE from its wild-type position at +15 [lane SVL-13A (WT)] to +23 (lane SVL-13A/23U) caused a partial shift in the site of cleavage from position +1 to +4, confirming our previous observations (26,27). Movement of the AAUAAA element in this SVL variant from its wild-type position at –13 to –8 resulted in efficient cleavage and a further downstream shift in the sites of cleavage to positions +6 and +9 (lane SVL-8A/23U). Since the construct containing the AAUAAA element at position –8 and the URE at position +15 resulted in inefficient cleavage (Fig. 1), these data demonstrate that the efficiency and location of cleavage are determined by the relative position of the AAUAAA and URE.

While the above data suggest that the AAUAAA and URE interact to determine parameters of cleavage, they do not address whether or not one element has more influence than the other. A recent study showed that while increasing the distance between the URE and the AAUAAA element shifted the site of cleavage to a more downstream location, movement of the URE closer to the AAUAAA element failed to shift the cleavage site to a more upstream location (26). This observation might be due to context effects or to a specific spacing requirement between the AAUAAA element and the site of cleavage. In order to decide between these two possibilities we tested the effect of alterations of the position of the URE in an SVL variant which contained the AAUAAA element at position –20. Movement of the URE from position +14 to position +20, +23 or +25 caused a progressive downstream shift in the site of cleavage as expected, due to the positional influence of the URE (data not shown). Movement of the URE from position +14 to position +9 or +6, on the other hand, had no effect on the site of cleavage, although the efficiency of cleavage was progressively decreased (data not shown). In

other words, cleavage was never detected closer than 14 bases or beyond 23 bases from the AAUAAA element in the context of this construct. In contrast, cleavage occurred within five bases of the URE, albeit inefficiently (26; data not shown). Cleavage was never observed to occur greater than 28 bases upstream from the URE in our SVL variants (26; data not shown).

These data argue that the relative position of the AAUAAA and U-rich elements is important in determining the site and efficiency of cleavage. Although both elements showed a maximal distance limit for cleavage site determination, only the AAUAAA element has the additional requirement of a minimal downstream distance for cleavage site placement.

Spatial requirements for the AAUAAA element and the URE can be applied to nearly three quarters of mammalian polyadenylation signals

We next tested whether the experimental observations described above could be extended to other polyadenylation signals. By searching GenBank as described in Materials and Methods we identified 131 individual mammalian polyadenylation signals for which sequence information was available for both the upstream and downstream regions. The site of cleavage/polyadenylation was inferred from the sequence of cDNA clones isolated from these genes. These polyadenylation signals were chosen without bias, other than the requirement that they all contain an AAUAAA element.

We first asked how many of these 131 polyadenylation signals contain an appropriately positioned URE. As seen in Figure 3, 102 out of 131 contained a four-out-of-five base URE. The observation that 77.9% of polyadenylation signals contained a URE argues strongly that the URE is the major downstream element of the polyadenylation signal. Furthermore, 88.2% of URE-containing polyadenylation signals had the URE located at the optimal range of 10–30 bases downstream of the cleavage site, which we have determined experimentally for the SVL polyadenylation signal (26). Only seven polyadenylation signals had a URE located closer than 10 bases from the cleavage site and only five had a URE located at positions between +30 and +39. However, it is possible that RNA secondary structures may bring a URE into an optimal position to mediate efficient cleavage in these polyadenylation signals. The possibility also exists that longer range secondary structures could bring a more distant URE into an appropriate position in those polyadenylation signals which did not contain an identifiable URE. These data, along with the observation that the URE serves as the binding site for the 64 kDa subunit of CstF (27), support the conclusion that a four-out-of-five base URE is the predominant downstream element of mammalian polyadenylation signals.

We next assessed the location of the AAUAAA element relative to the cleavage site in the set of polyadenylation signals. As shown in Figure 4, only one polyadenylation signal had its AAUAAA element located closer than 11 bases from the inferred site of cleavage, while 121 out of 131 signals contained the AAUAAA element at positions between 11 and 23 bases from the cleavage site. These data are entirely consistent with our experimental observations with position variants of the AAUAAA element of the SVL polyadenylation signal described above. We conclude that, with only a single exception, cleavage occurs no closer than 11 bases from the AAUAAA element. It is also very rare for cleavage to occur beyond 23 bases from the

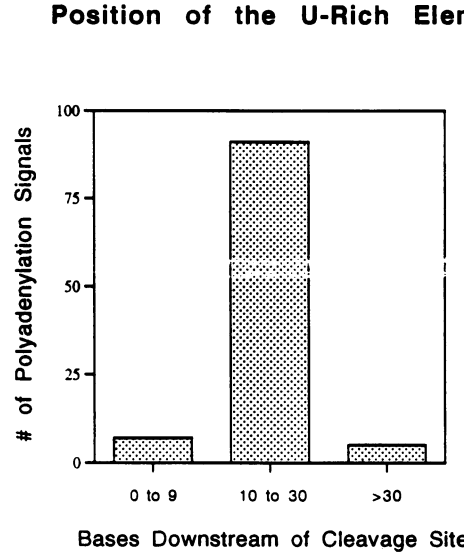

Figure 3. A computer survey confirms and generalizes the optimal position for the downstream URE in mammalian polyadenylation signals. The *x*-axis refers to the number of bases downstream of the cleavage site. The *y*-axis refers to the number of polyadenylation signals out of 131 surveyed.

AAUAAA element. The few exceptions to this maximal distance rule may be the result of RNA secondary structures which alter the physical distance between the AAUAAA element and the cleavage site.

Cleavage occurs preferentially at adenosine residues located within the region defined by the AAUAAA and URE

As discussed above, cleavage occurs within a region 11–23 bases downstream from the AAUAAA element. The site of cleavage within this 13 base region is influenced by the relative position of the downstream URE (Figs 1 and 2). The observation by Sheets *et al.* (10) that 59% of polyadenylation signals surveyed contained a C at position −1 and 71% contained an A at position +1 suggested that a CA dinucleotide at the site of cleavage may be an additional important recognition element for the cleavage machinery. In order to assess this possibility, we changed the CA dinucleotide located at positions −1 and +1 of the SVL polyadenylation signal to all 16 possible combinations of four nucleotides and assessed the effect of these changes on cleavage site selection.

Figure 5 shows representative data from 8 of the 16 variants we constructed. Several general observations can be made from these data. First, mutations at the +1 and/or −1 positions had only minor effects on the overall efficiency of cleavage. Second, although cleavage occurred at predominant sites, additional minor sites of cleavage over an ~6 base region were noted in our assays. This suggests that the cleavage domain identified by the AAUAAA and URE may span ~6 nt and that the cleavage machinery selects positions within this domain as predominant sites of cleavage. Third, in every mutant tested an adenosine residue was chosen as a predominant site of cleavage. Six out of 16 mutants also used a U residue as a major site of cleavage, while 2 of 16 also used a C residue. None of our variants chose a G residue for a cleavage

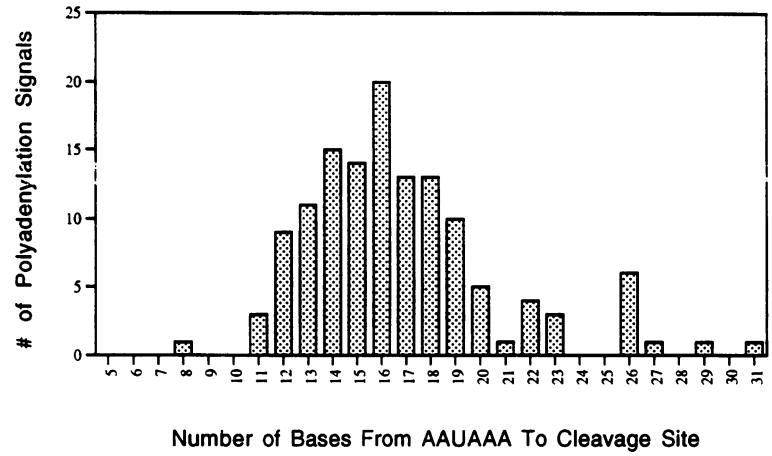

Figure 4. A computer survey confirms and generalizes an optimal position for the AAUAAA element in mammalian polyadenylation signals. The *x*-axis refers to the number of bases from the 3′ A residue of the AAUAAA element to the cleavage site. The *y*-axis refers to the number of polyadenylation signals out of 131 surveyed.

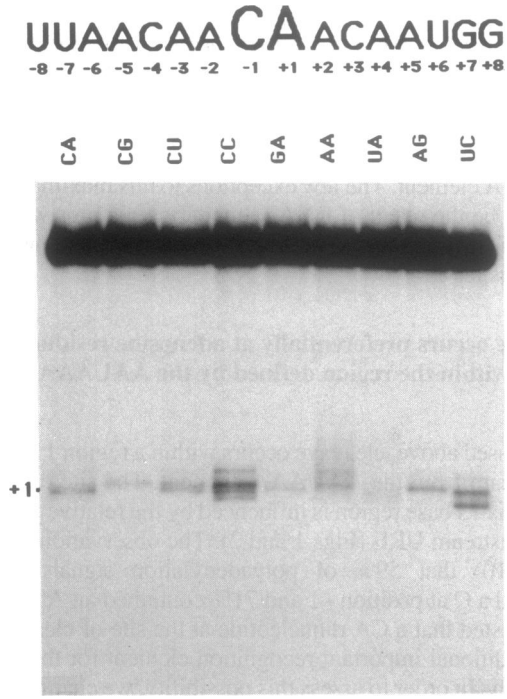

Figure 5. Nucleotides within the cleavage region defined by the AAUAAA and U-rich elements affect the site of cleavage. Variants of the SVL polyadenylation signal which contain the indicated nucleotides at the –1 and +1 positions were incubated in the *in vitro* cleavage system. RNA products were analyzed on a 5% polyacrylamide sequencing gel containing 8 M urea. The 5′ cleavage products at the wild-type position +1 are indicated. The wild-type sequence from position –8 to +8 relative to the cleavage site of the SVL polyadenylation signal is shown at the top.

site. These data suggest that cleavage site usage is influenced by nucleotide preference in the order A > U > C >> G.

Our mutational analysis found no evidence for a CA dinucleotide motif involved in cleavage site selection. Changing the C at position –1 to G, A or U still resulted in the use of the A at position +1 as a major site of cleavage (Fig. 5, lanes GA, AA and UA). Additional major cleavage sites which are used in some of these variants are the result of insertion of the more favorable cleavage nucleotides A or U for C at the –1 position. Changing the A at position +1 to G, U or C resulted in a partial or complete shift of the site of cleavage to the adjacent A residue at position +2 (Fig. 5, lanes CG, CU and CC). Some cleavage was also detected at +1 in the CU and CC variants. This cleavage site was presumably still active in these variants due to the substitution of an adenosine residue by the acceptable, but less preferred, U or C residues.

These data, therefore, are most consistent with a model in which a single nucleotide within a six base region is chosen by the cleavage machinery in the order of preference A > U > C >> G. All of the cleavage data obtained for the double mutants at the CA dinucleotide are consistent with this model (i.e. Fig. 5, lanes AG and UC).

DISCUSSION

In this report we have shown that the position of the AAUAAA element relative to a downstream URE influences the site and efficiency of cleavage in the SVL polyadenylation signal. Analysis of position variants in the SVL signal and a survey of 131 independent mammalian polyadenylation signals suggests that cleavage occurs at least 11 bases from the AAUAAA element, but no farther than 23 bases. Our analyses have also provided strong evidence that most polyadenylation signals contain a four-out-of-five base URE located within 30 bases downstream of the cleavage site. The optimal position for the URE is 10–30 bases downstream of the cleavage site. Finally, we have demonstrated that the precise site of cleavage within the region defined by the AAUAAA and URE is determined by a single nucleotide in the order of preference A > U > C >> G.

Based on these data we have developed a model for the functional relationship between the major *cis*-acting elements of the mammalian polyadenylation signal. As seen in Figure 6, cleavage occurs within a 13 base region located downstream from the AAUAAA element. The precise site of cleavage within this 13 base region is governed by three factors. First, it can be no closer than 11 bases downstream of the AAUAAA element. Second, the relative position of the URE and the AAUAAA element influences the site of cleavage. The URE is usually

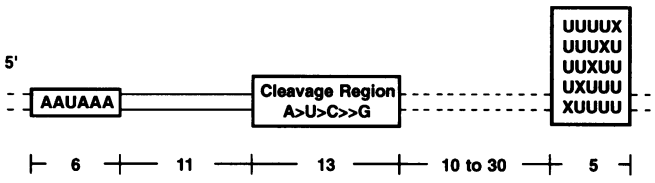

Figure 6. Functional relationships of the major *cis*-acting elements of the mammalian polyadenylation signal. The upstream AAUAAA element and the downstream URE are shown in the boxes. The sequence preferences for cleavage site selection are shown in the cleavage region box. The length of the major *cis*-acting elements, the cleavage region and the distance between these elements and the cleavage region (in nucleotides) are shown by the numbers on the bottom.

located 10–30 bases downstream of the cleavage site; if the URE is located in the downstream half of this defined range, cleavage will occur in the downstream portion of the 13 base cleavage region and vice versa. The relative positions of the AAUAAA and URE delimit an ~6 base domain in which cleavage will occur. Third, the local sequence within the cleavage region influences the precise site of cleavage. The order of sequence preference at the cleavage site is A > U > C >> G. We found no evidence for a dinucleotide as a recognition motif for the cleavage reaction.

The description in Figure 6 provides a working model for the major *cis*-acting elements of the polyadenylation signal which accounts for all of the previous experimental observations with URE-containing polyadenylation signals. Fitzgerald and Shenk, for example, showed that deletions between the AAUAAA element and the wild-type cleavage site shifted cleavage to locations 13 and 17 bases downstream of the AAUAAA element (9), precisely as predicted by our model. The model also successfully explains the results obtained with downstream deletion derivatives. For example, the spacing requirement for a 'GU-rich' downstream segment relative to the AAUAAA element defined by Cole and co-workers (15) reflects the position requirements for the URE described in our model, since the GU-rich segment of the HSV TK polyadenylation signal moved in these studies contained a URE. Second, the contradictory results observed by the Wickens (23) and Alwine (21,22) laboratories with the SVL polyadenylation signal can be explained by the observation that the downstream region of the signal has three UREs at position +15, +20 and +50. Deletions which remove only the URE at position +50 have no effect on the efficiency of 3′-end processing (35). Deletions which remove the UREs at positions +15/+20, which were previously used to argue for the functional significance of the region from +46 to +55 (22), actually move the URE at position +50 into a functional +11 context. Mutations downstream of +30 in the SVL polyadenylation signal which decrease the efficiency of polyadenylation do so by disrupting an auxiliary G-rich element (35). As a final example, the observation of Mason *et al.* (30) that deletions in the downstream region of the *Xenopus* β-globin polyadenylation signal caused a shift in the site of cleavage can be explained by the removal of the original URE and/or positioning of a new URE at position +10 to +17.

Approximately one quarter of mammalian polyadenylation signals do not appear to follow the model shown in Figure 6. There are several explanations for this subset of polyadenylation signals. First, RNA secondary structures may reposition UREs located at a distance to an appropriate location for mediating efficient 3′-end formation. Such secondary structures have been shown to realign a far upstream AAUAAA element to an appropriate location in the HTLV-1 polyadenylation signal (36) or in a synthetic polyadenylation signal (37). Second, alternative elements in addition to the URE may exist which mediate CstF positioning in these signals (16). Third, since the site of cleavage in most of the polyadenylation signals surveyed is inferred from cDNA sequence, rather than determined directly, it is possible that alternative polyadenylation sites may exist in these constructs.

In addition to the major *cis*-acting element AAUAAA and URE described in this study, auxiliary elements located both upstream and downstream of the cleavage site have been demonstrated to influence the efficiency of 3′-end processing. Mutation or movement (35,38) of these elements has not been shown, however, to influence the site of cleavage. Auxiliary elements are likely to promote efficient 3′-end processing by stabilizing the interaction between the general polyadenylation factors and the major *cis*-acting elements (35,38).

It has been shown that the 160 kDa protein of CPSF interacts with the major *cis*-acting element AAUAAA (12,13). The positional requirement for this element with regard to cleavage site determination implies that the CPSF complex covers ~11 bases downstream from the AAUAAA element. Consistent with this notion, *in vitro* polyadenylation of pre-cleaved synthetic substrates, which is dependent on CPSF, requires a minimum of eight bases downstream of the AAUAAA element (10). The three base discrepancy between the minimal distance requirement from the AAUAAA element for cleavage versus polyadenylation may reflect the spatial requirements of the different enzymes involved. It has also been demonstrated that the 64 kDa protein of CstF interacts in a stable fashion with the URE in the presence of CPSF (27,29,34). The positional requirements for both the AAUAAA and the U-rich elements described in this study are likely to reflect the spatial requirements for a stable interaction between CPSF and CstF. Furthermore, we hypothesize that CPSF and CstF form a scaffold and align the cleavage factors over the cleavage region (27). The precise site of cleavage is then influenced by the local nucleotide sequence within the cleavage region defined by the CPSF and CstF scaffold, following the order of sequence preference A > U > C >> G. The efficiency of cleavage at this location is governed by the stability of the assembly of these general processing factors on the RNA substrate. Complex stability is not only affected by the sequence and position of the major *cis*-acting elements, but can also be influenced by the presence of upstream or downstream auxiliary elements (35,38).

The work in this study provides a detailed description of the major elements of the polyadenylation signal, extending our understanding beyond the association of an upstream hexanucleotide AAUAAA element and a downstream GU- or U-rich region with 3′-end processing. The identification of a downstream consensus element, its positional relationship with the AAUAAA element and the sequence preference for cleavage site selection provide important information for a mechanistic understanding of 3′-end processing.

ACKNOWLEDGEMENTS

This work was supported by grant AI31165 from the National Institutes of Health to J.W. F.C. was supported by training grant CA09663-01. C.C.M. is an Associate of the Howard Hughes Medical Institute in the laboratory of Thomas Shenk. J.W. is a Pew Scholar in the Biomedical Sciences.

REFERENCES

1 Wahle,E. and Keller,W. (1992) *Annu. Rev. Biochem.*, **61**, 419–440.
2 Nevins,J.R. and Darnell,J.E. (1978) *Cell*, **15**, 1477–1493.
3 Moore,C.L. and Sharp,P.A. (1985) *Cell*, **41**, 845–855.
4 Peterson,M.L. (1994) *Mol. Cell. Biol.*, **14**, 7891–7898.
5 Connelly,S and Manley,J.L. (1988) *Genes Dev.*, **2**, 440–452.
6 Proudfoot,N.J. (1989) *Trends Biochem. Sci.*, **14**, 105–110.
7 Nesic,D. and Maquat,L.E. (1994) *Genes Dev.*, **8**, 363–375.
8 Niwa,M. and Berget,S.M. (1991) *Genes Dev.*, **5**, 2086–2095.
9 Fitzgerald,M. and Shenk,T. (1981) *Cell*, **24**, 251–260.
10 Sheets,M.D., Ogg,S.C. and Wickens,M.P. (1990) *Nucleic Acids Res.*, **18**, 5799–5805.
11 Wilusz,J., Pettine,S.M. and Shenk,T. (1989) *Nucleic Acids Res.*, **17**, 3899–3908.
12 Keller,W., Bienroth,S., Lang,K.M. and Christofori,G. (1991) *EMBO J.*, **10**, 4241–4249.
13 Murthy,K.G.K. and Manley,J.L. (1992) *J. Biol. Chem.*, **267**, 14804–14811.
14 Gil,A. and Proudfoot,N.J. (1987) *Cell*, **49**, 399–406.
15 Heath,C.V., Denome,R.M. and Cole,C.N. (1990) *J. Biol. Chem.*, **265**, 9098–9104.
16 McDevitt,M.A., Hart,R.P., Wong,W.W. and Nevins,J.R. (1986) *EMBO J.*, **5**, 2907–2913.
17 McDevitt,M.A., Imperiale,M.J., Ali,H. and Nevins,J.R. (1984) *Cell*, **37**, 993–999.
18 McLauchlan,J.D., Gaffney,J.L., Whitton,J.L. and Clements,J.B. (1985) *Nucleic Acids Res.*, **13**, 1347–1368.
19 Ryner,L.C. and Manley,J.L. (1987) *Mol. Cell. Biol.*, **7**, 495–503.
20 Ryner,L.C., Takagaki,Y. and Manley,J.L. (1989) *Mol. Cell. Biol.*, **9**, 1759–1771.
21 Sadofsky,M. and Alwine,J.C. (1984) *Mol. Cell. Biol.*, **4**, 1460–1468.
22 Sadofsky,M., Connelly,S., Manley,J.L. and Alwine,J.C. (1985) *Mol. Cell. Biol.*, **5**, 2713–2719.
23 Zarkower,D. and Wickens,M. (1988) *J. Biol. Chem.*, **263**, 5780–5788.
24 Berget,S.M. (1984) *Nature*, **309**, 179–182.
25 Renan,M.J. (1987) *Gene*, **60**, 245–254.
26 Chou,Z.-F., Chen,F. and Wilusz,J. (1994) *Nucleic Acids Res.*, **22**, 2525–2531.
27 MacDonald,C.C., Wilusz,J. and Shenk,T. (1994) *Mol. Cell. Biol.*, **14**, 6647–6654.
28 Takagaki,Y., Manley,J.L., MacDonald,C.C., Wilusz,J. and Shenk,T. (1990) *Genes Dev.*, **4**, 2112–2120.
29 Weiss,E.A., Gilmartin,G.M. and Nevins,J.R. (1991) *EMBO J.*, **10**, 215–219.
30 Mason,P.J., Elkington,J.A., Lloyd,M.M., Jones,M.B. and Williams,J.G. (1986) *Cell*, **46**, 263–270.
31 Wilusz,J. and Shenk,T. (1988) *Cell*, **52**, 221–228.
32 Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
33 Wilusz,J. and Shenk,T. (1990) *Mol. Cell. Biol.*, **10**, 6397–6407.
34 Wilusz,J., Shenk,T., Takagaki,Y. and Manley,J.L. (1990) *Mol. Cell. Biol.*, **10**, 1244–1248.
35 Bagga,P.S., Ford,L.P., Chen,F. and Wilusz,J. (1995) *Nucleic Acids Res.*, **23**, 1625–1631.
36 Ahmed,Y.F., Gilmartin,G.M., Hanly,S.M., Nevins,J.R. and Greene,W.C. (1991) *Cell*, **64**, 727–737.
37 Brown,P.H., Tiley,L.S. and Cullen,B.R. (1991) *Genes Dev.*, **5**, 1277–1284.
38 Gilmartin,G.M., Fleming,E.S., Oetjen,J. and Graveley,B.R. (1995) *Genes Dev.*, **9**, 72–83.